



Spatio-temporal uncertainty in Spatial Decision Support Systems: A case study of changing land availability for bioenergy crops in Mozambique

Judith Anne Verstegen^{a,*}, Derek Karssenber^a, Floor van der Hilst^b, André Faaij^b

^a Department of Physical Geography, Faculty of Geosciences, Utrecht University, 3508 TC Utrecht, The Netherlands

^b Copernicus Institute for Sustainable Development and Innovation, Utrecht University, The Netherlands

ARTICLE INFO

Article history:

Received 11 May 2011

Received in revised form 25 August 2011

Accepted 29 August 2011

Available online 4 October 2011

Keywords:

Spatial Decision Support Systems

Uncertainty

Spatial modeling

Visualization

Land use change

Bioenergy

ABSTRACT

Spatial Decision Support Systems (SDSSs) often include models that can be used to assess the impact of possible decisions. These models usually simulate complex spatio-temporal phenomena, with input variables and parameters that are often hard to measure. The resulting model uncertainty is, however, rarely communicated to the user, so that current SDSSs yield clear, but therefore sometimes deceptively precise outputs. Inclusion of uncertainty in SDSSs requires modeling methods to calculate uncertainty and tools to visualize indicators of uncertainty that can be understood by its users, having mostly limited knowledge of spatial statistics. This research makes an important step towards a solution of this issue. It illustrates the construction of the PCRaster Land Use Change model (PLUC) that integrates simulation, uncertainty analysis and visualization. It uses the PCRaster Python framework, which comprises both a spatio-temporal modeling framework and a Monte Carlo analysis framework that together produce stochastic maps, which can be visualized with the Aguila software, included in the PCRaster Python distribution package. This is illustrated by a case study for Mozambique in which it is evaluated where bioenergy crops can be cultivated without endangering nature areas and food production now and in the near future, when population and food intake per capita will increase and thus arable land and pasture areas are likely to expand. It is shown how the uncertainty of the input variables and model parameters effects the model outcomes. Evaluation of spatio-temporal uncertainty patterns has provided new insights in the modeled land use system about, e.g., the shape of concentric rings around cities. In addition, the visualization modes give uncertainty information in a comprehensible way for users without specialist knowledge of statistics, for example by means of confidence intervals for potential bioenergy crop yields. The coupling of spatio-temporal uncertainty analysis to the simulation model is considered a major step forward in the exposure of uncertainty in SDSSs.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Spatial Decision Support Systems (SDSSs) are interactive, computer-based systems that include simulation models and visualization tools designed to assess the impact of possible decisions (Geertman & Stillwell, 2004). With SDSSs, planners can investigate the effects of different scenarios, and explore intervention possibilities by adjusting model inputs in a user interface accessible to users that do not have expert knowledge of modeling theory and technology. The models in SDSSs usually simulate change over time of a spatial phenomenon or, more likely, a number of spatial phenomena that interact with each other. These dynamic processes and interactions tend to be complex and are rarely fully understood (Manson, 2007). As simulation models are simplifications of open, complex systems, model output errors are inherent

as a result of the debatable choice and conceptualization of relevant sub-processes, uncertainty in model parameters and input variables, and the discretization of information. These errors propagate through the model because the state of the modeled system at a certain moment in time is a function of its state in the past. This generates uncertainty in model outputs.

Whereas scientists are familiar with the concept of uncertainty and methods to quantify it (Brown & Heuvelink, 2007; Chen, Wood, Linstead, & Maltby, 2011; Goodchild, 2004; Heuvelink, 1998), SDSS users tend to seek certainty and deterministic solutions (Bradshaw & Borchers, 2000). In other words, they demand practical models with clear and unambiguous results to facilitate decision making. The result of this desire for clarity and simplicity is that SDSSs tend to underestimate, if not ignore, uncertainty (Foody, 2003). They thus yield clear, but therefore sometimes deceptively precise outputs.

Decisions based on misinterpreted or erroneous model output can be costly due to the irreversibility of such decisions.

* Corresponding author. Tel.: +31 (0)30 2535513.

E-mail address: J.A.Verstegen@uu.nl (J.A. Verstegen).

Uncertainty thus needs to be communicated clearly. Therefore we claim, together with others (e.g., Aerts, Goodchild, & Heuvelink, 2003; Foody, 2003; Ivanovic & Freer, 2009; Ma, Arentze, Borgers, & Timmermans, 2007; Manson, 2007; Oreskes, Shrader-Frechette, & Belitz, 1994), that instead of obscuring uncertainty users should be made more aware of uncertainty in SDSSs. Difficulties in including and communicating uncertainty in SDSSs are: (1) the concepts and measures of uncertainty can be somewhat difficult to grasp for users without specialist knowledge of statistics, (2) uncertainty is input-dependent, (3) uncertainty varies over space, time and aggregation level, and (4) software packages that can integrate spatio-temporal modeling, uncertainty analysis and visualization are rare.

The first difficulty arises from the fact that it is unlikely that the average user of an SDSS has skills in handling uncertainty at a comparable level as the modelers themselves (Foody, 2003). Therefore, modelers should aid their end users by providing intuitive and insightful indicators of uncertainty and straightforward tools to visualize these (Aerts, Goodchild, et al., 2003).

The second problem is that the model uncertainty cannot be calculated on forehand by the modeler, as output uncertainty depends on model inputs and parameters. Currently, uncertainty in simulation models is sometimes assessed by providing a (static) map of output uncertainty of the final time step of a standard run of the simulation model (e.g., Brown, Page, Riolo, Zellner, & Rand, 2005; Chang, Parvathinathan, & Breeden, 2008; Eckhardt, Breuer, & Frede, 2003). Clearly, this does not suffice for an SDSS, in which investigating the effects different model settings is the main goal, so that inputs and parameters are altered frequently. For that reason, uncertainty analysis should be automatically calculated by the model itself, every time it is run with different settings.

The third difficulty is that uncertainty varies over space and time, because complex systems behave non-linearly. This makes that providing an uncertainty map of the final time step only is not sufficient when it comes to complex, non-linear systems (Ligmann-Zielinska & Sun, 2010). So, an uncertainty map is needed for each time step, in order to allow *iterative uncertainty analysis* (Manson, 2007), as for example demonstrated by Gorsevski, Gessler, Boll, Elliot, and Foltz (2006). In addition, it has been shown by others (Hiemstra & Karssenber, in press; Kok, Farrow, Veldkamp, & Verburg, 2001; Pontius & Spencer, 2005) that uncertainty is highly dependent on the level of spatial aggregation. Usually, uncertainty becomes lower at a coarser scale, because rearrangements in the landscape at locations in close proximity cancel each other out when they are aggregated. This is highly relevant in SDSSs, as their end users operate at different managerial levels (e.g., town, district, province, country), related to spatial scale levels (local, regional, national). To be able to aid these different end users, it should be possible in an SDSS to assess uncertainty at different levels of aggregation. Ideally, inclusion of such methods should be possible without too much additional work on the side of the model developer.

The final problem is that most software packages are either dedicated to model development, e.g., Stella (2010) and NetLogo (2010), or to uncertainty analysis (for a package overview see Goovaerts, 2010), or to visualization, e.g., ArcGIS (ESRI, 2010). Using such packages to construct an uncertainty-inclusive SDSS would require a complex coupling mechanism. Also, existing visualization tools do not explicitly support methods to visualize uncertain spatio-temporal data. A possible solution to this problem is the PCRaster model construction framework (Karssenber, Schmitz, Salamon, de Jong, & Bierkens, 2010; PCRaster, 2010), which offers a combined interface for spatio-temporal modeling and uncertainty analysis and includes a visualization tool for stochastic data in its distribution package.

The objective of this paper is to construct an SDSS that integrates simulation, iterative uncertainty analysis, and visualization to facilitate end users at different managerial levels to take uncertainty into account in decision making. This is illustrated by a case study of bioenergy-crop potentials in Mozambique. Although some studies have been conducted to assess the area of potentially available land for bioenergy crops in Mozambique (Batidzirai, Faaij, & Smeets, 2006; Watson, 2011), none of these studies was carried out in both a temporally dynamic and spatially explicit way. The PCRaster Land Use Change model (PLUC) is developed to evaluate where bioenergy crops can be cultivated without entering into competition with other important land uses from a economic or sustainability point of view, now and in the near future when population and food intake per capita and thus arable land is likely to increase. We show that PLUC allows stochastic model inputs and produces interactive visualizations of forecast uncertainty, in space and time, at a range of spatial aggregation levels. These visualizations can be used by decision makers to evaluate possible locations and potential yields for bioenergy crops.

The next section of this paper describes the concepts and methods of the PCRaster Python framework, outlines how the framework is applied to construct the land use change model for Mozambique, explains the error models of the different stochastic inputs, and illustrates the mode of implementation. The results section shows different visualization modes of uncertainty indicators and their potential usage, drawing on the outputs of PLUC for the Mozambique case study. The final section discusses the advantages and shortcomings of the uncertainty-inclusive simulation model.

2. Methodology

2.1. Software framework

Although other approaches exist to include uncertainty in a model, such as fuzzy logic (e.g., Nguyen, de Kok, & Titus, 2007; Robinson, 2003), stochastic modeling has the advantage that it has a strong root in mathematics. In stochastic modeling, a model input is defined by a probability distribution of all possible values. This uncertainty is propagated through the model using a numerical solution scheme. Monte Carlo simulation is such a solution scheme, which is attractive because of its general applicability and ease of implementation (Aerts, Goodchild, et al., 2003). It involves running the model a large number of times, each time drawing a realization from the input probability distribution(s). For spatial models this results in different spatial patterns for the different model realizations, i.e. model runs or samples.

The PCRaster model construction framework (Karssenber et al., 2010; PCRaster, 2010) facilitates this integration of spatio-temporal modeling and uncertainty analysis through the PCRaster Python library (Karssenber, de Jong, & van der Kwast, 2007). This library provides a large set of spatio-temporal functions on raster maps, embedded in the Python language (Python, 2010). Both a spatio-temporal modeling framework and a Monte Carlo analysis framework are present as a Python class. These classes include methods to write the simulation results and uncertainty indicators to disk as maps, which can be visualized with the Aguila software

Table 1
Modeling schedule of the PCRaster Python framework.

1	for each MC sample:
2	for each time step:
3	solve system state equation
4	compute summary statistics

(Pebesma, de Jong, & Briggs, 2007), included in the PCRaster Python distribution package. To allow construction of a spatio-temporal model that permits stochastic inputs and assessment of the resulting uncertainty, three main methods are provided by the framework that together form the schedule in Table 1 (Karssenbergh & De Jong, 2006). Firstly, there is a loop for evaluation of the spatio-temporal process itself (line 2). Herein, the modeler can program the equations that represent the change of the system state over a time step (line 3). The framework provides for this purpose a number of functions particularly designed for spatial and stochastic operations. Secondly, a loop over this spatio-temporal model is performed to generate the Monte Carlo samples (line 1). Finally, summary statistics are computed over all Monte Carlo samples (line 4), representing the uncertainty in the model output within a time step or over the whole simulated period of time.

The individual Monte Carlo sample results and summary statistics are written to disk with a function from the PCRaster Python library. This function uses rules for file names defined by the modeling framework, so that they can directly be visualized with the *Aguila* software that recognizes these name conventions. Temporal deterministic and stochastic data can be viewed by animation or toggling through time, of which the last technique was considered helpful by decision makers (experts as well as novices) for visualizing uncertain data in a study of Aerts, Clarke, and Keuper (2003). In addition, stochastic outputs are visualized as maps or plots of mean, standard deviation, confidence interval, exceedance probability, or cumulative probability distribution of a variable. These visualizations are interactive, which allows users to explore the data (Karssenbergh et al., 2010; Pebesma et al., 2007).

2.2. Land use change model

The potential to employ the PCRaster Python framework for an SDSS including uncertainty is illustrated by the construction of a land use change model, meant to aid in evaluating where bioenergy crop plantations can be allocated. Until now, bioenergy potentials were mostly assessed in a spatially aggregated (e.g., Hoogwijk, Faaij, Eickhout, De Vries, & Turkenburg, 2005) or temporally static (e.g., van der Hilst et al., 2010) way. So, in this study a both spatially explicit and temporally dynamic model is created for Mozambique. The case is relevant, because Mozambique is considered promising for bioenergy crop production by its vast amounts of available land (Smeets, Faaij, Lewandowski, & Turkenburg, 2007), favorable environmental conditions for cultivation (Batidzirai et al., 2006), and relatively low productivity of current agriculture (Arndt, Benfica, Tarp, Thurlow, & Uaiene, 2010), which offers potential for improvement. However, over the past decade the area of forests and woodlands has decreased substantially due to an increase in cultivated areas (Jansen, Bagnoli, & Focacci, 2008). Now as well as in the future, cultivation of bioenergy crops should not add to that effect and not endanger other important land uses, either from an economic, e.g., food crops and livestock, or from a sustainability point of view, e.g., conservation areas (Haberl, Beringer, Bhattacharya, Erb, & Hoogwijk, 2010). The population is expected to increase and its diet is expected to change as well, which induces further shifts in land use. The direction and extent of this shift depends on the agricultural and livestock productivity, which is expected to improve. The trends herein are derived from literature (e.g., FAO, 2003; INE, 2003), fieldwork, and meetings with national and local authorities. More background on the modeled processes and used data is provided in the twin-publication of this paper (van der Hilst, Versteegen, Karssenbergh, & Faaij, submitted for publication). This section focuses on the set-up of the model.

The main model component is the state transition function representing the change in spatially distributed land use over a time step. Many models of land use dynamics have been constructed be-

fore (see for overviews Agarwal, Green, Grove, Evans, & Schweik, 2002; Parker, Manson, Janssen, Hoffmann, & Deadman, 2003; Verburg, Schot, Dijst, & Veldkamp, 2004). Some focus on one specific land use conversion, such as urbanization (Batty, 2005; Ligtenberg, Beulens, Kettenis, Bregt, & Wachowicz, 2009), but more often several land uses in the area compete for new locations (Lei, Pijanowski, Alexandridis, & Olson, 2005; Verburg & Overmars, 2009). We adopt the latter approach, in which we focus on active change in agricultural land use types and forest as the future distribution of land reserved for these land uses is the main issue in the view of potential locations for bioenergy crop plantations. By *active change*, we mean that expansion or contraction of the total area of this land use is explicitly steered by certain drivers. Other land use types on the land use map can change passively, by expansion or contraction of an active land use type.

The land use change is steered by two factors: (1) the demand of the population for food, non-food crops (e.g., cotton and tobacco) and wood, and (2) the growth rate of yield, defined by agricultural and livestock productivity. The conceptual model (Fig. 1) includes three loops. The first one loops over time and coincides with line 2 in Table 1. The other two belong to the actual state transition function (line 3 in Table 1); one loops over all active land use types and the other checks iteratively whether this land use type should expand, contract or has met its demand. The actual location of the expansion or contraction of the land use types is determined by suitability factors, like distance to cities and transport networks, current land use in the neighborhood, and location-specific yield due to characteristics of the soil and climate. Areas occupied by other economically important land uses, physically constrained areas, and protected land uses are excluded and the remaining land is potentially available for bioenergy crops. But the bioenergy crops are not included as a land use type, which means they are not allocated. The model is explained in more detail in the following.

An important step in model implementation is the choice of the support to represent processes (Bierkens, Finke, & de Willigen, 2000; Hengl, 2006; Pan, Roth, Yu, & Doluschitz Reiner, 2010). The support refers to the size in the spatial domain (i.e. spatial discretisation), and the temporal domain (i.e. time step duration) over which processes are considered homogeneous. Model inputs and parameters need to be representative for the support used (c.f., Bierkens et al., 2000). The following criteria were used to select a suitable support. (1) The scale at which the model output is required in order to answer the end user's questions (Evans & Kelley, 2004), also called *policy scale* (Bierkens et al., 2000). In our case, policy makers at different spatial scale levels are involved, ranging from national to local scale. Information is needed on changes in land use over the coming decades. (2) The *process scale*, i.e. the scale of natural variability of the studied process (Blöschl, 1999). We do not aim to study land use changes at parcel level; the smallest transformation of interest is at the level of a small community of farmers. (3) *Observational data availability*. If the input data do not match the resolution of the model, upscaling or downscaling could be applied, but one should be careful to do so, as spatio-temporal probability distributions alter with changing resolution (Bierkens et al., 2000). Thus it is preferable to choose a support that matches the support of available observational data. (4) *Processing power*, as calculation time increases exponentially with the number of cells (Hengl, 2006). Because PLUC is designed as an SDSS, it should be usable on a desktop pc. The Monte Carlo simulations should be performable in a reasonable amount of time and should result in a manageable amount of data. As the system studied constitutes of a number of subsystems that operate at different scales, while data availability differs between subsystems, we use three levels of spatial support: (1) country level, for data that is only available for the country as a whole, like economic trends, (2) land use type level, for crop-dependent but location-independent variables, like maximum possible product yield, and (3) cell level,

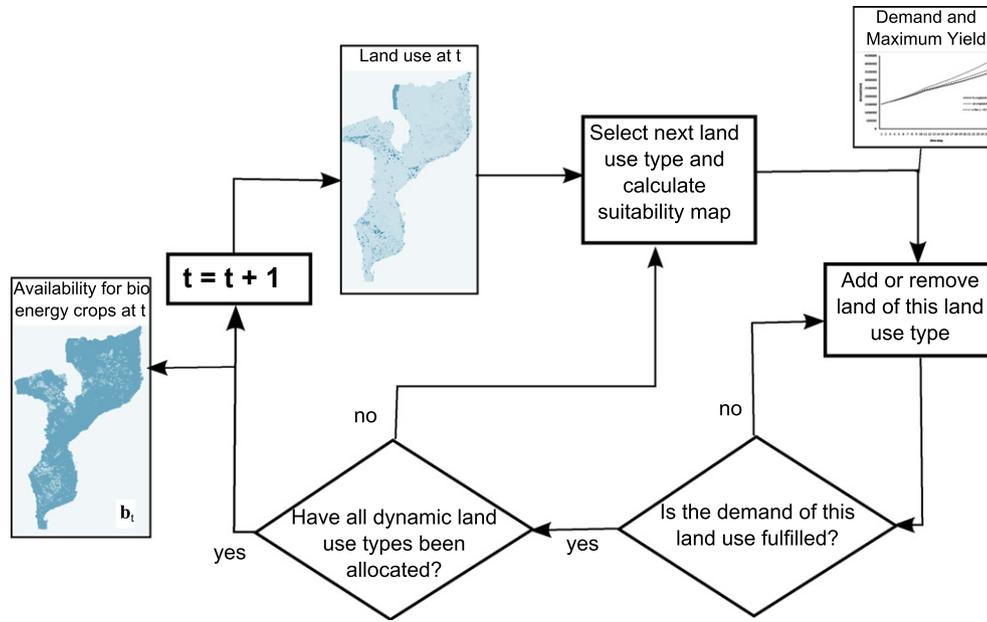


Fig. 1. Conceptual model of land use change.

for location-specific information, like population density. These three levels are implemented using gridded maps with a cell size of 1 km². This means that technically, all information is discretized into cells of 1 km², but in fact the three abovementioned levels of spatial support are used in input data and process descriptions. The model uses a time step (Δt) of one year, with time step $t = 1, 2, \dots, T$. Most equations are evaluated separately for each of the N land use types, with $n = 1, 2, \dots, N$. The model is run for 25 years and the following dynamic land use types are defined: cropland, mosaic cropland–pasture (grazed grassland), mosaic cropland–grassland (not grazed), pasture, forest. However, any cell size, model period, and number and type of land uses could be defined by the user.

The demand $d_{n,t}$ (kg year⁻¹) for products from land use type n at time step t is:

$$d_{n,t} = p_t * i_{n,t} * r_t, \quad \text{for each } n \text{ in each } t \quad (1)$$

In Eq. (1), p_t denotes the number of inhabitants in the country at t , intake $i_{n,t}$ (kg caput⁻¹ year⁻¹) specifies the demand per capita of products from land use type n at t , and the self-sufficiency ratio $r_t(-)$ is the extent to which the food demands are met by domestic supply at t .

Potential yield $p_{n,t}$ (kg km⁻² year⁻¹) is the yield of products from land use type n at t if the cell would be occupied by that land use type:

$$p_{n,t} = m_{n,t} * f_n, \quad \text{for each } n \text{ in each } t \quad (2)$$

In Eq. (2), $m_{n,t}$ is the maximum possible product yield (kg km⁻² year⁻¹) of products from land use type n at t , which can increase through time due to technological improvements in the agricultural or livestock sector, i.e. increased productivity (FAO, 2003). The location-specific variable $f_n \in [0, 1]$ is the actual fraction of this yield that can be reached in a cell, depending on factors like soil type, climate, and water availability. Note that a bold font indicates that a variable is a spatial field, i.e. information at cell level.

The total product yield of a land use type is calculated using the land use map at t . First, a spatial field of the current yield $c_{n,t}$ (kg km⁻² year⁻¹) of land use type n at t is constructed that contains the value of $p_{n,t}$ for cells that are currently occupied by type n at t

and zero for cells occupied by other land use types. The total product yield $y_{n,t}$ (kg) of this land use type is therefore:

$$y_{n,t} = \sum (c_{n,t} * a), \quad \text{for each } n \text{ in each } t \quad (3)$$

In Eq. (3), the summation sign indicates summation over the whole spatial field and a is the cell size in km² (1 km² in this study).

To determine where a certain land use expands or contracts, every land use type is assigned a number of suitability factors. In total, nine suitability factors have been implemented in PLUC. The number and kind of suitability factors differ per land use type. Table 2 shows which suitability factors were implemented for every dynamic land use type in the case study of Mozambique and which weights were assigned to them (see Eq. (4) below). For a detailed explanation of why these factors and weights were chosen, the reader is referred to van der Hilst et al. (submitted for publication).

The spatial autocorrelation suitability factor (1) assumes that land uses attract land uses of the same or a related type (e.g., cropland is related to mosaic cropland–pasture and mosaic cropland–grassland), i.e. that related land uses tend to cluster. It determines the area of land use type n or types that are related to n in the neighborhood of a cell with land use type n . The size of the neighborhood is determined by the window length l (m), i.e. the length of the square window around the center cell, and the area occupied by the same or related land use types is calculated relative to the size of the window. This suitability factor thus introduces positive feedback loop into the model: if a land use type is allocated at a certain location, the suitability value in the neighborhood of that location increases for this land use type, so that more of this land use might be allocated in the area in the next time step.

The suitability factors for distance to roads (2), water (3), cities (4) and edge of the plot (8) determine suitability based on the shortest Euclidean distance to the object under consideration, i.e. roads, water, cities, or edges of plots, i.e. spatially connected areas of a uniform land use type. For crops and pasture typically a location close to roads and cities is preferred in order to minimize transport costs, and close to water for irrigation. Wood is preferably harvested at the edge of the forest, because this makes harvesting easier. The distance suitability functions (2, 3, 4 and 8) all have two parameters: relation type between distance and suitability (linear, exponential or inversely proportional) and range r_n (m).

Table 2
Weights per suitability factor i per land use type n . The suitability factors concern: spatial autocorrelation (1), distance to roads (2), water (3), and cities (4), yield (5), population density (6), livestock density (7), distance to plot edge (8), and conversion elasticity (9).

Land use type (n)	Suitability factor (i)								
	1	2	3	4	5	6	7	8	9
Cropland	0.20	0.10	0.10	0.10	0.20	0.20	0	0	0.10
Cropland–grassland	0.20	0.10	0.10	0.10	0.20	0.20	0	0	0.10
Cropland–pasture	0.20	0.10	0.10	0.10	0.15	0.15	0.10	0	0.10
Pasture	0.30	0.05	0.15	0.05	0.1	0.05	0.20	0	0.10
Forest	0.25	0	0	0.20	0.05	0.30	0	0.20	0

The range indicates the maximum distance of effect of the feature, e.g., the maximum distance of effect of a road on the land use type pasture is set at 5 km, based on fieldwork, expert knowledge and literature (e.g., Jansen et al., 2008). Cells at a distance of more than the range have a suitability value of zero for this suitability factor.

The yield fraction (5), population density (6) and livestock density (7) suitability factors relate to the fraction of the maximum that can be found in a cell (see Eq. (2)), where the maximum refers to maximum yield ($m_{n,t}$), maximum population density and maximum livestock density in factor 5, 6, and 7, respectively. Per land use type the direction (increasing or decreasing) and relation type between fraction and suitability (linear, exponential or inversely proportional) can be indicated, like in the distance suitability functions, e.g., cropland is preferably located on cells with a high yield, but wood might preferably be harvested from cells with a low yield, i.e. biomass, because harvesting is easier in sparse forest.

The current land use (9) suitability factor indicates the compliance of a certain land use type to be transformed into the land use type that implements the suitability factor. This factor is sometimes referred to as conversion elasticity (Verburg & Overmars, 2009), e.g., it is more preferable for pasture to be placed on a cell that is currently defined as ‘abandoned’ than on a cell that is currently defined as ‘cropland’, as the second case involves a greater loss of economic value. Note that most spatial fields resulting from the suitability factors ($\mathbf{u}_{i,n,t}$, see Eq. (4) below) remain the same over time, as the location of the features it relates to, e.g., roads, does not change over time in the model. This is the case for the factors 2, 3, 4, 5, 6, and 7. However, the spatial fields of suitability factors related to dynamic land use types (1, 8, and 9) do change over time and thus establish feedback loops in the land use system.

For every land use type a total suitability map $\mathbf{s}_{n,t} \in [0, 1]$, indicating the aggregated appropriateness of a given location for land use n at time step t , is computed from its suitability factors:

$$\mathbf{s}_{n,t} = \sum_{i=1}^9 (w_{i,n} * \mathbf{u}_{i,n,t}), \quad \text{for each } n \text{ in each } t$$

$$\text{with } \sum_{i=1}^9 (w_{i,n}) = 1 \quad (4)$$

In Eq. (4), $\mathbf{u}_{i,n,t} \in [0, 1]$ is the spatial field resulting from suitability factor i for land use type n at time step t , and $w_{i,n} \in [0, 1]$ is the weight of suitability factor i for land use type n (see Table 2).

Now, all information is available to allocate the land uses. The land use types have a certain hierarchy determined mainly by their economic importance. In Africa, the current trend is that agricultural intensification takes place on land that is now used as mosaic cropland–pasture or mosaic cropland–grassland. This extensive land use becomes less common and is moved to less fertile grounds, while areas even less fertile, like mountain areas and forests, come into use for grazing of livestock (Lambin et al., 2001). Therefore, we assume the following order of allocation for the dynamic land uses: cropland, mosaic cropland–pasture, mosaic cropland–grassland, pasture, forest. The allocation schema can be written in pseudo-code as in Table 3. When the land use type expands, it allocates new cells of this type at locations with the

Table 3

Pseudo-code for land use allocation procedure. Allocation of each land use type n in each time step t proceeds until yield fulfills demand.

```

1  if  $\mathbf{d}_{n,t} > \mathbf{y}_{n,t}$ :
2    while  $\mathbf{d}_{n,t} > \mathbf{y}_{n,t}$ :
3      convert cell with max ( $\mathbf{s}_{n,t}$ ) to  $n$ 
4      update  $\mathbf{y}_{n,t}$ 
5  else if  $\mathbf{d}_{n,t} < \mathbf{y}_{n,t}$ :
6    while  $\mathbf{d}_{n,t} < \mathbf{y}_{n,t}$ :
7      convert cell with min ( $\mathbf{s}_{n,t}$ ) to 99
8      update  $\mathbf{y}_{n,t}$ 
9  else:
10 do nothing

```

highest suitability, i.e. $\max(\mathbf{s}_{n,t})$, and when it contracts it removes cells of this type with the lowest suitability, i.e. $\min(\mathbf{s}_{n,t})$. Cells are converted to or removed from this land use until the total yield $\mathbf{y}_{n,t}$ equals the total demand $\mathbf{d}_{n,t}$. In Table 3 the number 99 in line 7 refers to the land use category ‘abandoned’. For land use type forest the class ‘abandoned’ is named ‘deforested’ in order to be able to distinguish cleared forest from deserted agricultural land on the resulting land use map.

When allocation of one land use type is finished, allocation of the next type is performed, with the restriction that it cannot convert cells with a land use type that has already been allocated in that time step. Deforested areas become forest again when they are left fallow for 10 years. The regenerated forest can be harvested once more to fulfill the wood demand.

At the end of each time step, when the land use map has changed according to the demands of the different land use types, it is determined which cells are potentially available for bioenergy crops. This is done by excluding all areas occupied by crops, pasture, steep slopes (calculated from the digital elevation model), roads, water, cities, forest concession areas, community areas, and nature reservation areas. This results in a Boolean map (\mathbf{b}_t) where cells are available (True) or unavailable (False) for bioenergy crops at time step t . Although the bioenergy crops are not allocated on the land use map, they do have their own maximum possible product yield m_t and yield fraction map \mathbf{f} , so the bioenergy crop yield per location and in total for the available area \mathbf{b}_t can be calculated using Eqs. (2) and (3). These results can be used to assess the available area and yield on provincial and national level, in order to take the influence of spatial aggregation into account.

2.3. Error models

The various projections of population growth, diet change and technological improvements in the agricultural sector differ significantly (Arndt et al., 2010; FAO, 2003; UNDP, 2008), so applying one of these datasets deterministically to quantify drivers presumably ignores a large input error. Also, a number of model parameters are uncertain as they can only be estimated by expert knowledge, because extensive model calibration datasets are cur-

Table 4

Stochastic variables of the land use change model and their data type, error model and standard deviation (σ), if applicable. For explanation of error models, see main text.

Stochastic variable	Data type	Error model	σ
Window length (l)	Single value	Normal	3 km
Range (r_n)	Single value	Uniform	–
Elevation (\mathbf{h})	Spatial field	Normal	1 m
Yield fraction (\mathbf{f}_n)	Spatial field	Relative normal	0.2
Population density	Spatial field	Relative normal	0.1
Livestock density	Spatial field	Relative normal	0.1
Maximum yield ($\mathbf{m}_{n,t}$)	Time series	Relative normal	0.1
Demand ($\mathbf{d}_{n,t}$)	Time series	Uniform	–

rently not available. PLUC takes each of these input errors into account in calculating the forecast uncertainty. Model drivers and parameters that are uncertain are defined here as lumped or spatially distributed stochastic variables. The variables can be divided into three groups according to their data type: single value, spatial field and time series (Table 4).

Two stochastic variables represent a single value: the window length l used in suitability factor 1 and the range r_n , used in suitability factors 2, 3, 4, and 8. For window length l a normal error model is used:

$$l = \mu_l + Z_l * \sigma_l \quad \text{with } Z_l \sim N(0, 1) \quad (5)$$

In Eq. (5), μ_l is the mean of l , i.e. the value that would be used in a deterministic run. In our case study a value of 3 km is used for μ_l , as this means that only the direct neighbors are taken into account in our 1×1 km raster. In Eq. (5), σ_l is the standard deviation, which is set to 1 km. It should be noted that l can attain values such that the window cuts through cells. This is not a problem as the suitability factor calculates the area in the window occupied by attracting land use types, not the number of cells.

For the range r_n an error model with a uniform distribution is used:

$$r_n = \sqrt{a} + Z_r * 2 * \mu_{r,n} \quad \text{with } Z_r \sim U(0, 1) \quad (6)$$

In Eq. (6), a is the cell size in km^2 . This means that realizations of the range vary between the cell length and twice the mean value that is used in a deterministic run. This lower limit is used, because land use cannot be allocated on the feature under consideration, e.g., a road.

Four stochastic input variables represent a spatial field, i.e. a raster map of values. For surface elevation \mathbf{h} a normal error is used:

$$\mathbf{h} = \boldsymbol{\mu}_h + \mathbf{Z}_h * \sigma_h \quad \text{with } \mathbf{Z}_h \sim \mathbf{N}(\mathbf{0}, \mathbf{1}) \quad (7)$$

In Eq. (7), $\boldsymbol{\mu}_h$ is the original elevation map and σ_h is the standard deviation, for which a value of 1 m is used. Note that the normal error \mathbf{Z}_h is a spatial field, which means that a separate value is drawn for each cell. The other three stochastic spatial fields are the yield fraction, population density and livestock density. For all three a relative normal error model is used. This means that the error (in this case a normal error) is higher for higher mean values. For example, the yield fraction \mathbf{f}_n is defined as:

$$\mathbf{f}_n = \boldsymbol{\mu}_f + \mathbf{Z}_f * \sigma_f * \mu_f \quad \text{with } \mathbf{Z}_f \sim \mathbf{N}(\mathbf{0}, 1) \quad (8)$$

In Eq. (8), $\boldsymbol{\mu}_f$ is the original yield fraction map and σ_f is the standard deviation, for which a value of 0.2 is used. The population density and livestock density both have a standard deviation of 0.1. The standard deviation of the yield fraction is higher, because we found several different spatial data sets of the yield that were distinctively different, which indicates a large input error.

Finally, two stochastic time series are used. The first is the maximum yield $m_{n,t}$, to which a relative normal error model is assigned, similar to the one explained in Eq. (8):

$$m_{n,t} = \mu_{m,n,t} + Z_m * \sigma_m * \mu_{m,n,t} \quad \text{with } Z_m \sim N(0, 1) \quad (9)$$

In Eq. (9), $\mu_{m,n,t}$ is the mean of the of the maximum yield of land use type n at time step t , i.e. the expected value obtained from observation data or expert knowledge, and σ_m is the standard deviation, for which a value of 0.1 is used. The second time series is for the demand $d_{n,t}$. It uses an error model based on a uniform distribution between the upper and lower limit of the attribute:

$$d_{n,t} = l_t + Z_d * (u_{n,t} - l_{n,t}) \quad \text{with } Z_d \sim U(0, 1) \quad (10)$$

In Eq. (10), $l_{n,t}$ and $u_{n,t}$ are the lower limit and upper limit of the demand of land use type n at time step t . The limits of the demand $d_{n,t}$ are determined by upper and lower limits of population p_t , intake $i_{n,t}$ and self-sufficiency ratio r_t (Eq. (1)) predicted by the FAO (2003) and expert knowledge. Note that, although the resulting variables from Eqs. (9) and (10), $m_{n,t}$ and $d_{n,t}$, change over time, the stochastic variables Z_d and Z_m are drawn once, at the start of the simulation. These variables are used on all land use types, because they simulate the effect of the rate of increase in population, which cannot be different for the different land use types in the same model run.

2.4. Implementation

The schedule of PLUC is given in Table 5. The PCRaster Python framework consists of four main methods that represent the scheme in Table 1: the `premcloop` (line 6 in Table 5) is evaluated only once, the `initial` (line 21) is evaluated once for each realization, the `dynamic` (line 33) is evaluated once for each time step in each realization, and the `postmcloop` (line 54) calculates the descriptive statistics over the Monte Carlo samples. Some variables have the prefix `self`, as they are defined as member variables to allow usage over the four different methods.

When the `LandUseChangeModel` is initiated it calls PCRaster Python's `DynamicModel` (line 3) and `MonteCarloModel` (line 4). Among other things, these allow retrieving time steps (line 64) and Monte Carlo samples (line 65). Next, the `premcloop` (line 6) imports input maps in lines 7–16. A separate file, defined by the model builder, `parameters.py`, defines all non-spatial inputs. This is done to prevent that the end user has to make changes in the main model scheme. All its variables and parameters are imported in lines 17–20.

The `initial` method (line 21) is used to define initial or temporally constant stochastic variables for which a realization is drawn for each Monte Carlo sample. An example of an initial variable is the environment, i.e. the land use map, initiated in line 22. Examples of temporally constant stochastic variables are Z_d and Z_m used in Eqs. (10) and (9) for the calculation of demand $d_{n,t}$ and maximum yield $m_{n,t}$. For Z_d a value is drawn from a uniform distribution between 0 and 1 with the PCRaster Python `mapuniform()` function (lines 23). For the Z_m , a value from a normal distribution, from the PCRaster Python `mapnormal()` function, is multiplied by the standard deviation defined by the user (line 24). Next, the class `LandUse`, defined by the model builder, is instantiated (line 25). It is used to keep track of the changing land use map. This class has a method to calculate all other realizations for the stochastic variables in Table 4 (line 26). This `LandUse` class also instantiates N objects (i.e. one for each land use type) of the class `LandUseType` (line 27) that handle the land use type specific tasks, like computing suitability maps and allocating land (Eqs. (2)–(4) and Table 3). These methods are implemented with functions from the PCRaster Python library, including point,

Table 5
Main scheme of land use change model. Three dots and a discontinuity in the line numbering indicate omitted sections.

```

1  class LandUseChangeModel(DynamicModel, MonteCarloModel):
2  def __init__(self):
3      DynamicModel.__init__(self)
4      MonteCarloModel.__init__(self)
5      setclone('landuse')
6
7  def premcloop(self):
8      self.initialEnvironment = self.readmap('landuse')
...
17     self.landUseList = parameters.getLandUseList()
...
21     def initial(self):
22         self.environment = self.initialEnvironment
23         self.demandStoch = mapuniform()
24         self.maxYieldStoch = mapnormal() * self.sdYield
25         self.landUse = LandUse(self.landUseList, self.environment)
26         self.landUse.drawRealizationsParams(self.stochParams)
27         self.landUse.createLandUseTypeObjects(self.relatedTypeDict
, self.suitFactorDict, self.weightDict, self.varDict)
...
33     def dynamic(self):
34         demandUp = timeinputscalar('deUp.tss', self.environment)
35         demandLow = timeinputscalar('deLow.tss', self.environment)
36         demandDiff = (demandUp - demandLow)
37         demand = demandDiff * self.demandStoch + demandLow
...
42         self.landUse.calculateSuitabilityMaps()
43         self.landUse.allocate(maxYield, demand)
44         self.landUse.growForest()
45         self.environment = self.landUse.getEnvironment()
46         self.report(self.environment, 'landuse')
47         eu, euPr, euTo = self.landUse.getBioPotential(self.bioNoGo, \
self.provinces)
...
56     def postmcloop(self):
57         name = ['eu', 'euPr', 'euTo']
58         mcaveragevariance(name, self.sampleNumbers(), \
self.timeSteps())
59         name = ['eY', 'eYPr', 'eYTo']
60         percent = [0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95]
61         mcpercentiles(name, percent, self.sampleNumbers(), \
self.timeSteps())
62     nrOfTimeSteps = parameters.getNrTimeSteps()
63     nrOfSamples = parameters.getNrSamples()
64     myModel = LandUseChangeModel()
65     dynamicModel = DynamicFramework(myModel, nrOfTimeSteps)
66     mcModel = MonteCarloFramework(dynamicModel, nrOfSamples)
67     mcModel.run()

```

neighborhood, and global operations (Burrough & McDonnell, 1998). In lines 28–32 other initial actions are taken, like computing a map of the distance to roads, needed by the `LandUseType` class for calculation of the suitability maps.

In the `dynamic` method (line 33), the temporal components of the model are evaluated. Demand is defined as a stochastic input variable by two time series per land use type (lines 34 and 35), as explained in the error model section. The variable `demandStoch`, drawn in line 23, is used as the position between the upper and lower bound, $l_{n,t}$ and $u_{n,t}$, to calculate the realization for $d_{n,t}$ (Eq. (10)) (lines 37 and 38). In lines 38–41 the maximum yield $m_{n,t}$ is determined with the variable `maxYieldStoch` from line 24 (see Eq. (9)) in a similar way. Next, the suitability maps are calculated (line 42), and the allocation procedure, explained in Table 3, is called (line 43 in Table 5). The total land use map is updated (line 45) and saved to disk with the PCRaster Python function `report()` (line 46) that creates file extensions recognizable for the Aguil software. Next, it is determined which space is left over for bioenergy crops. In lines 47 and 48 it is determined where the bioenergy crop eucalyptus could be allocated and what its potential yield is on a cell-basis (`eu`), aggregated per province (`euPr`), and in total for the whole country (`euTo`).

In the `postmcloop`, the PCRaster Python function `mcaveragevariance` (line 58) calculates mean, variance and standard error of the files defined in line 57, for all time steps in all samples. The function `mcpercentiles` (line 61) computes the percentiles specified in the list in line 60 for a new list of files (line 59). All these estimators of uncertainty are automatically saved to disk. Note that these last two methods from the PCRaster Python library are the part of the model that calculate uncertainty. So, addition of just these two methods can turn any model into an uncertainty-inclusive model, given that it employs stochastic variables.

3. Results

The model was run in deterministic mode and in Monte Carlo mode using 500 samples to project land use change in Mozambique from 2005–2030. The Monte Carlo run implements the error models defined in Table 4, while the deterministic run takes the mean of each of these variables. This section focuses on added value of the uncertainty analysis, an in-depth discussion of the simulated land use patterns and potential bioenergy crop areas is provided in van der Hilst et al. (submitted for publication).

Fig. 2 shows how some land use maps resulting from a deterministic run of the model can be spatially and temporally explored using the Aguil visualization tool. Table 6 quantitatively summarizes the areas occupied by the dynamic land use types. Table 6 and Fig. 2 give an impression of the expansions and contractions of the different land use types. The spider-web-like pattern of deforestation (dark purple¹) is a result of wood harvesting near roads. Roads namely form both the edge the forest plot, so that harvesting is easier, and a transportation possibility. The bands of deforestation become broader over time, but stabilize more or less in the last two time steps shown, as forest then starts to regrow at cells that were emptied at the beginning of the simulation. Cropland (red) expands, mainly around cities at the cost of (mosaic) cropland-pasture (light green). This can be seen as agricultural intensification and specialization. The land use type cropland-pasture that represents the extensive self subsistence farming practices including extensive cultivation of crops and grazing of livestock on the same plot, is relocated to areas that have been cleared by the harvest of wood. Areas of (mosaic) cropland-grassland (yellow) expand as well, but more at the outer border (away from cities) of existing cropland-grassland and cropland-pasture plots, because they have less economic value and are thus allocated further away from population centers (see close up in the lower right panel in Fig. 2). Pasture intensification, i.e. conversion from cropland-pasture to 'pure' pasture (light purple), takes place primarily in the North-East of Mozambique, where the largest concentrations of livestock, in this case goats, are present. For pasture it is less essential than for crops that they are located close to a market place, as animals are self-transporting and are taken to the market less frequently (von Thünen, 1966), so they are located even further away from the city than cropland-grassland. In this way concentric rings of land use evolve, as predicted by, e.g., William Alonso's Bid Rent Theory (Alonso, 1964) and the Von Thünen model (von Thünen, 1966). A lot of information can be derived from this deterministic output, but it gives no information about the certainty of the observations, i.e. how general and how certain are the observed patterns?

To study this, Aguil allows visualization of different Monte Carlo runs in linked views, which make comparison easy. Fig. 3 shows the variation within three out of the 500 generated samples by providing a close up of land use in 2022 in the area around the city Nampula. It can be seen that the overall pattern of concentric

¹ For interpretation of color in Figs. 1–7, the reader is referred to the web version of this article.

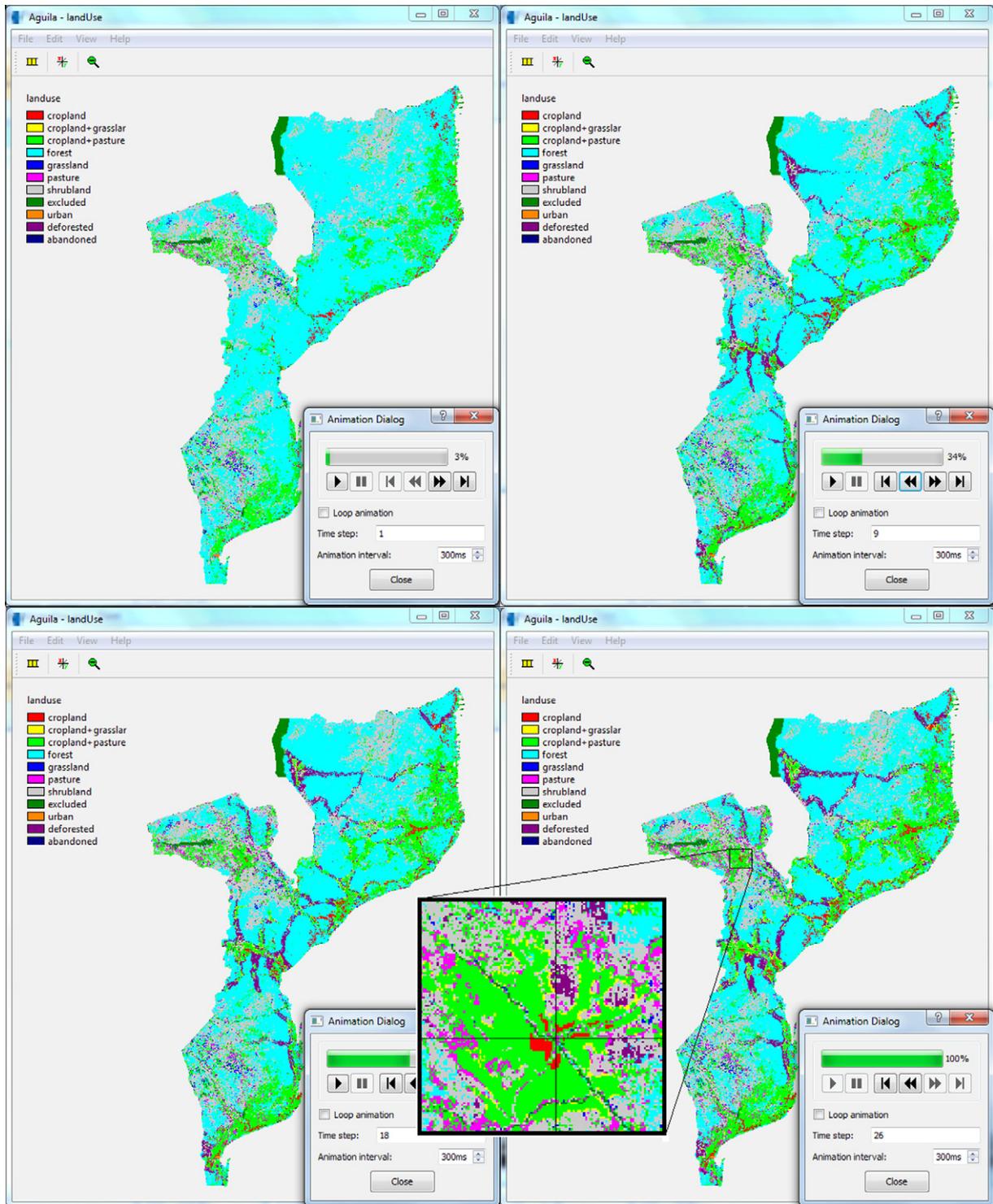


Fig. 2. Screenshot of Aguilá resulting from a deterministic model run showing land use in Mozambique in 2005 (time step 1, top left), 2013 (time step 9, top right), 2022 (time step 18, bottom left) and 2030 (time step 26, bottom right) with a close up around the city Tete.

Table 6
Simulated areas (km²) for the dynamic land use types in 2005, 2013, 2022 and 2030.

Year	Cropland	Mosaic cropland–pasture	Mosaic cropland–grassland	Pasture	Forest
2005	9382	123,737	2	7287	466,205
2013	11,639	135,188	5230	6835	418,654
2022	14,137	146,212	13,809	6179	391,540
2030	16,483	155,297	23,405	5618	368,805

rings is the same for all three realizations, but some differences are visible (see, e.g., the values of selected cell in the cursor window). One is that it is evident that the center and right image have some abandoned (dark blue) cells, while the left one has none. This can be explained by the fact that the values for demand ($d_{n,t}$) and maximum yield ($m_{n,t}$) are sampled separately. Cells that are abandoned on the center and right image are most often classified as cropland-pasture (light green) in the left one, so we take this land use type as an example for the explanation. It is given as a model input that both the demand and maximum yield of cropland-pasture increase over the modeled period. The growth rate however, differs per realization, depending on the values of Z_d and Z_m . As a result, three different situations are possible. The first situation applies for the center and right image: the demand for cropland pasture has increased, but agricultural and livestock productivity have increased so much that in the area currently occupied by cropland-pasture too much yield is generated. As a result the land use type cropland-pasture contracts in order to balance yield with demand. The second situation is that the agricultural and livestock productivity have not increased enough to counteract the effect of increasing demand, so the land use type expands. The last situation is an equilibrium, in which the increases in demand and maximum yield are in balance, so that no expansion or contraction is necessary. It cannot be derived directly from the left map in Fig. 3 which of the last two situations has occurred there, but it can be assumed that it is the second situation, expansion, as an exact equilibrium situation is very improbable. The described process is complicated by the fact that land of a certain type can be taken by another type during the simulation, so the current yield has to be updated constantly to check which situation is at hand. This stresses the fact that model output cannot be directly related to the input uncertainties due to the numerous non-linearities in the model.

Another observation that can be made in Fig. 3 is the shape of the concentric ‘rings’ around Nampula. If we focus on cropland (red), the center image shows circular clustering around the city, while in the right image it has a more star-like shape, with lumps

around the four roads connecting to Nampula. The left image has a shape somewhat in between. This difference is an effect of two suitability factors: distance to roads (suitability factor 2) and distance to cities (suitability factor 4). For cropland both factors are used (see Table 2), with the same weight $w_{i,t}$, which means they have equal effects on the total suitability for cropland. For both suitability factors a separate realization is made for the stochastic parameter Z_r , which determines the range r_n , i.e. maximum distance of effect. Fig. 4 illustrates how these two range parameters effect the sum of the suitability factors 2 and 4. The figure shows that a smaller value for the range of roads results in more clustering around roads and consequently star-shaped concentric ‘rings’ (Fig. 4b), while a smaller value for the range of cities results in

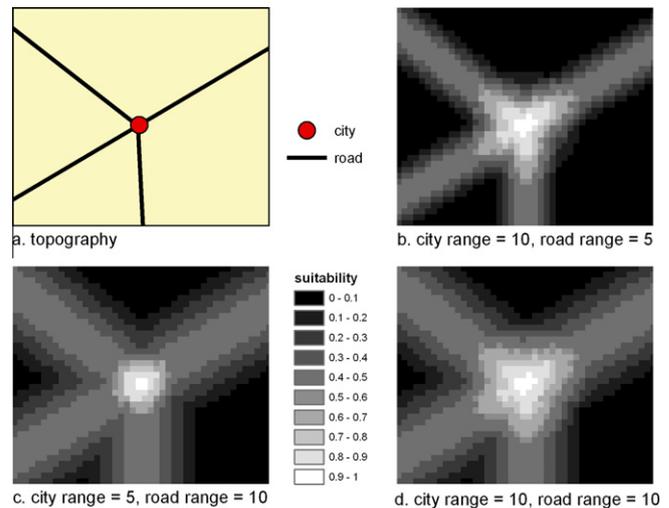


Fig. 4. Artificial dataset showing the effect of the range parameters of city and road on the total suitability and consequently on the shape of the concentric ‘rings’.

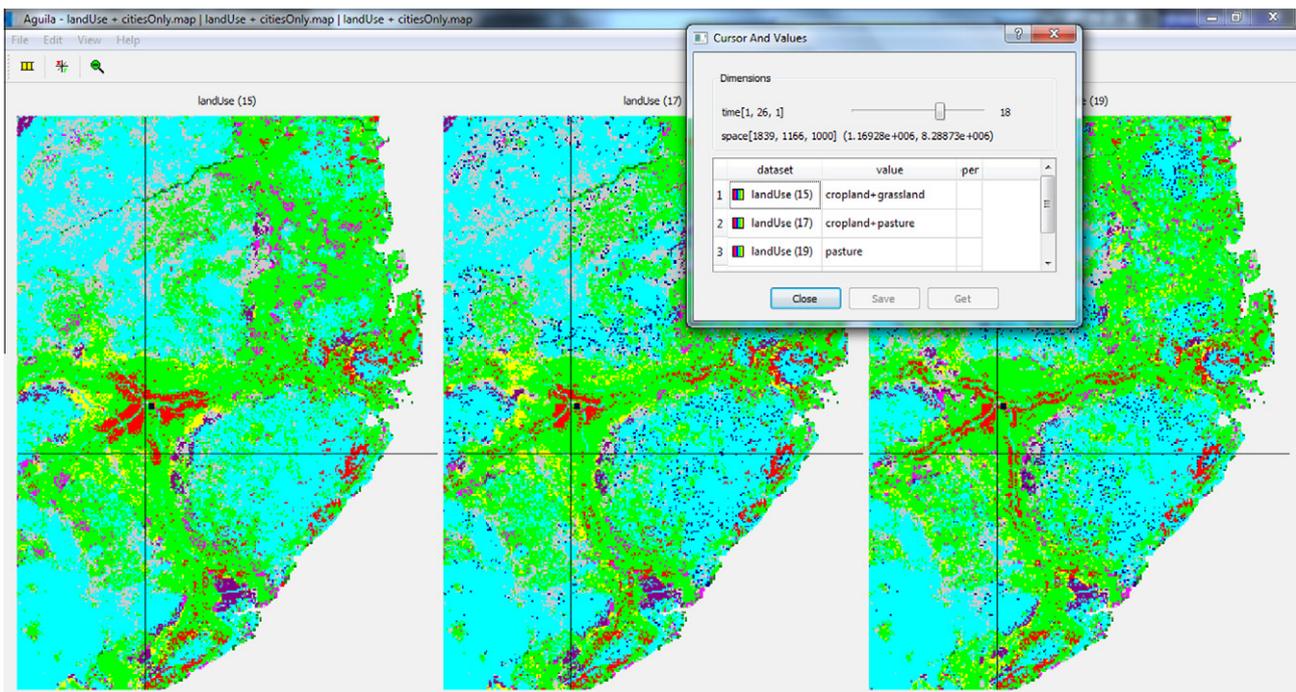


Fig. 3. Three different realizations of land use in 2022 (time step 18) zoomed in around the city Nampula, indicated in black, and the cursor window (top right) showing the land use type of the selected cell (cross in map views). The legend is the same as the one given in Fig. 2.

more clustering around cities and consequently circular concentric 'rings' (Fig. 4c). Similar values for the two ranges results in a shape in between (Fig. 4d).

The total area covered by the agricultural land use types and their spatial distribution determine what land is available for bioenergy crops. For this purpose, it is not very convenient to look at all realizations separately. Therefore, a result from the summary statistics over all 500 Monte Carlo samples is used (see line 58 in Table 5). Fig. 5 shows the probability that the bioenergy crop eucalyptus can be cultivated in a cell at a certain point in time without interfering with other important land uses. A value 1 indicates that a cell is certainly available, i.e. it was available in all realizations, a value 0 that it is certainly unavailable, i.e. it was available in none of the realizations, and any value in between indicates uncertainty in availability. In 2005 some land is available for eucalyptus around Tete, but in 2013 a ring has formed around the city that is certainly unavailable. This can be explained by the formation of concentric rings of agriculture around the city that may not be disturbed by

the cultivation of eucalyptus. The zone around Tete that is unavailable for eucalyptus becomes larger over time. It can be seen that the edges of the ring have a value somewhere between 0 and 1. This is because the size and shape of the concentric rings differ between the samples, so that cells at the edges of the concentric ring are in some samples occupied by agricultural land use types in the considered year and in some samples not. In the first case they are not available for eucalyptus and the second they are. This uncertainty ring 'moves' away from the city and roads through time, as the concentric rings around the city grow.

The planning of bioenergy crop plantations is not only of interest to local decision makers, but also of concern at higher managerial levels, e.g., province and country level. At these levels, the possible yield of eucalyptus per province and for the whole country at a certain point in the future and the uncertainty in these predictions are relevant. Fig. 6 shows the variance in potential eucalyptus yield at three different aggregation levels. The yield is calculated per km² at all three levels for comparability reasons. The variance

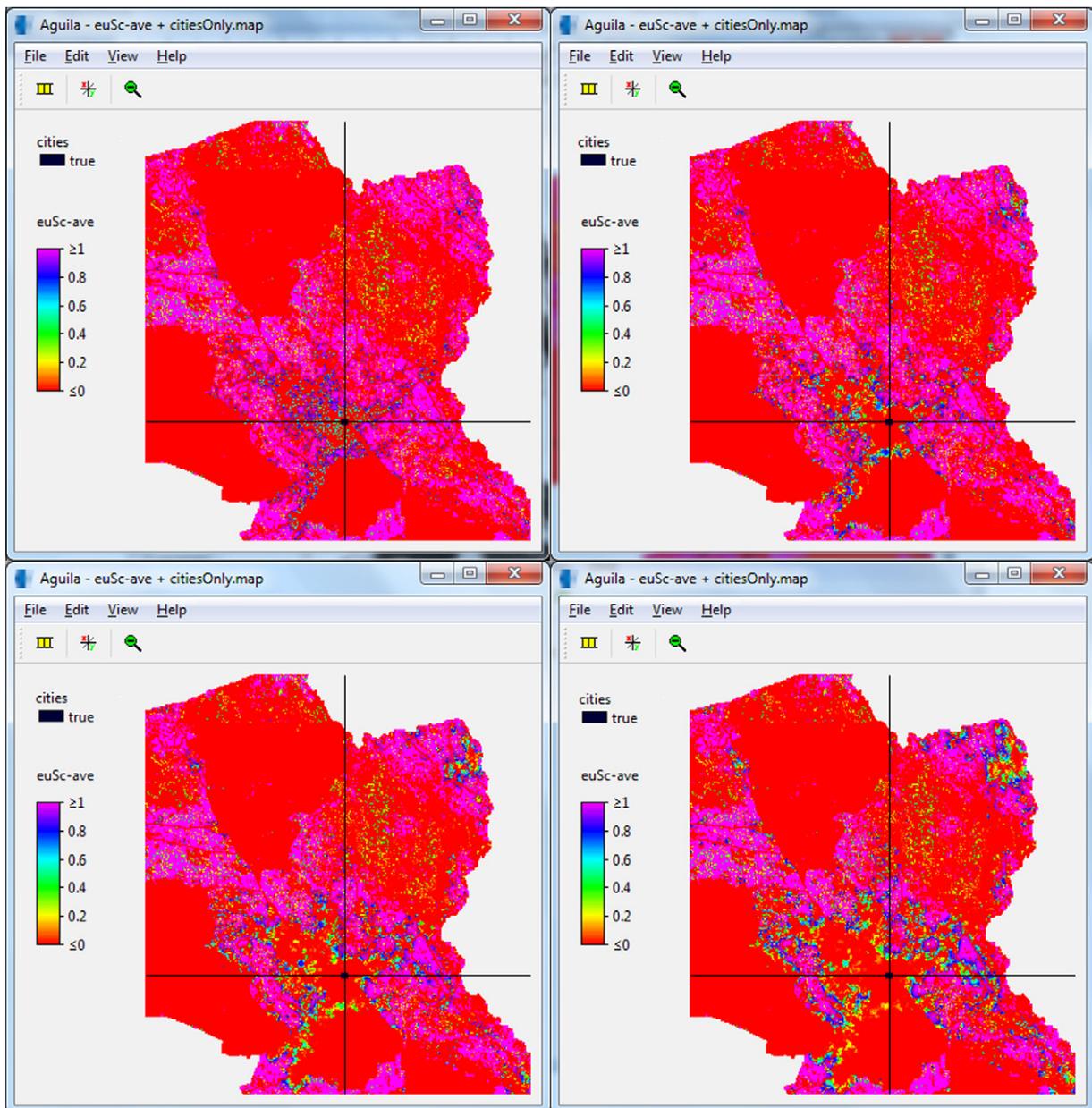


Fig. 5. Probability that a cell is available for bioenergy (eucalyptus) in 2005 (time step 1, top left), 2013 (time step 9, top right), 2022 (time step 18, bottom left) and 2030 (time step 26, bottom right) zoomed in around the city Tete indicated in black.

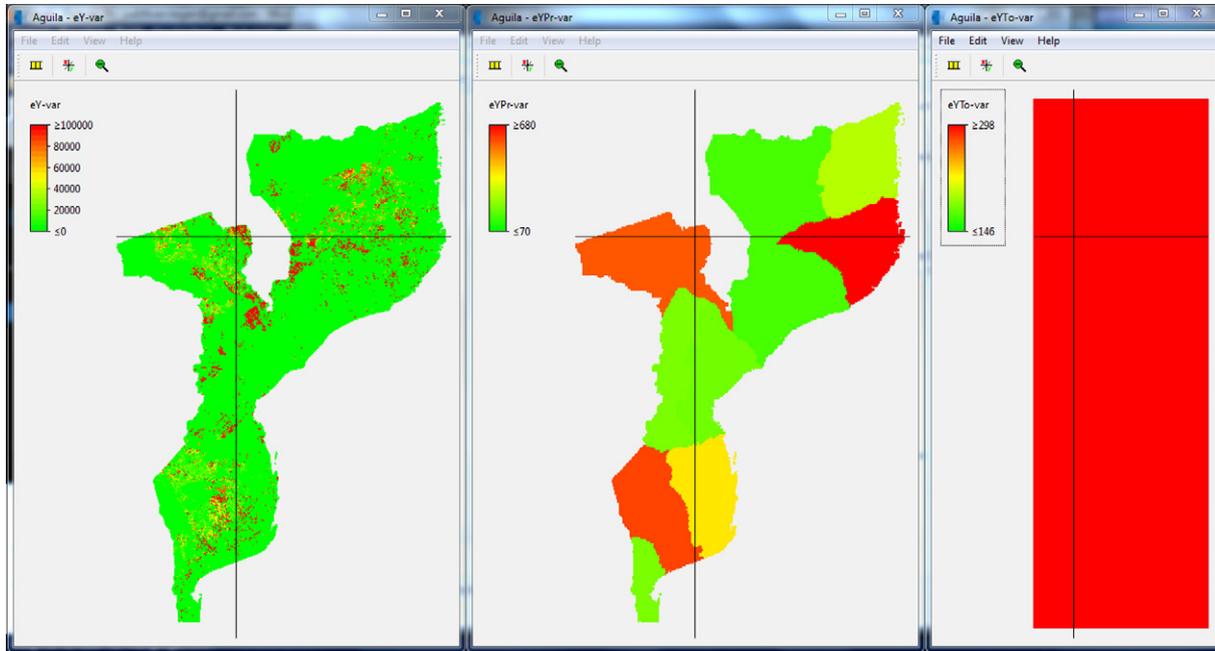


Fig. 6. Variance in yield (10^{-5} kg km⁻² year⁻¹)² per cell (left), per province (middle) and for the country as a whole (right) in 2030.

in yield is a combined result of cell availability (Fig. 5), yield fraction and maximum yield of eucalyptus as explained in the following. Most cells in Fig. 6 have a variance of zero, because they have an availability probability of zero, or because the soil is infertile (yield fraction of zero). Some cells have a variance slightly above zero; these cells have an availability probability of one, but their yield differs because of the stochastic parameters in yield fraction (z_f) and maximum yield (z_m). Finally, there are some cells with very high variances; these cells are sometimes unavailable, and then have a yield of zero, and sometime available, and then have a yield

dependent on the stochastic parameters yield fraction and maximum yield. This generates very large variances. It is evident that the maximum variance (maximum value of the value scale bar shown on the left side of the three maps) becomes much lower when scaling from cell level (1×10^5 (kg km⁻² year⁻¹)²) to province level (680 (kg km⁻² year⁻¹)²) to country level (298 (kg km⁻² year⁻¹)²). This is because local differences between samples are levelled out at higher aggregation levels.

Other uncertainty information that can be used are the calculated percentiles (see line 61 in Table 5). The lower part of Fig. 7

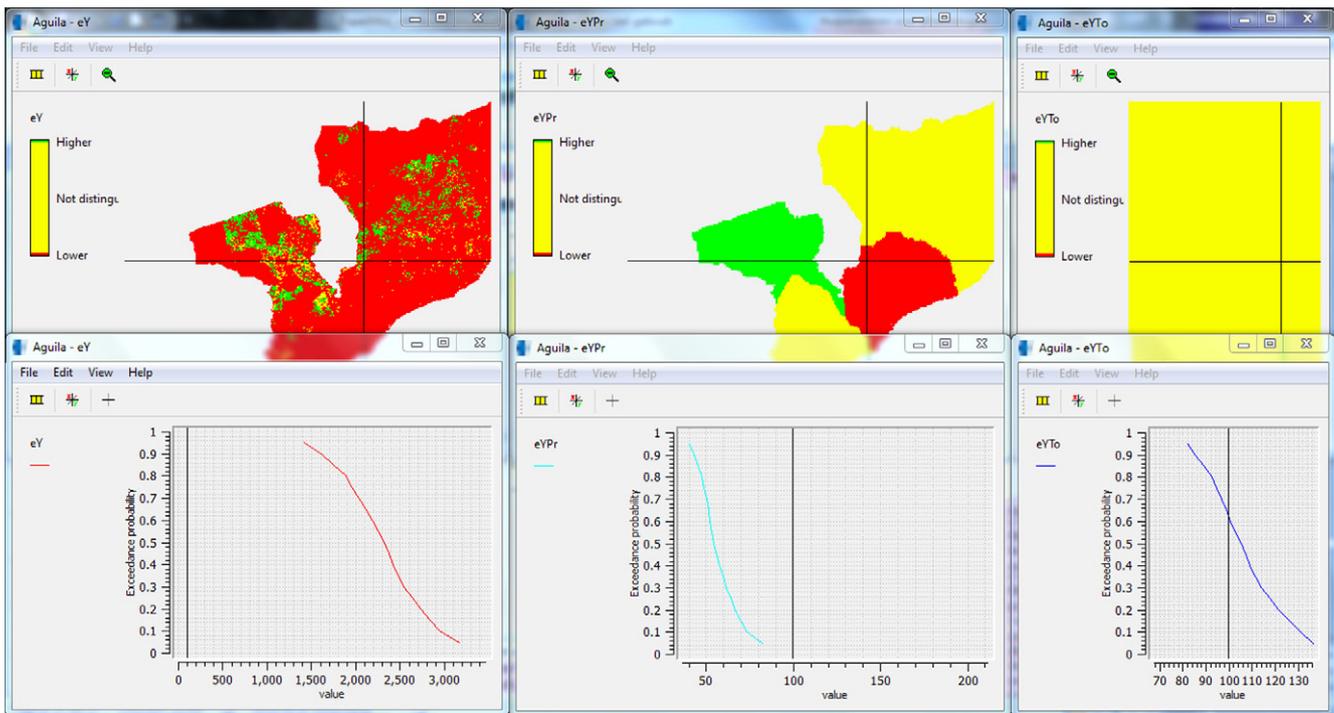


Fig. 7. Top: 95% confidence interval for a yield above a threshold of 100 kg km⁻² year⁻¹ in 2030. Higher = certainly above, Lower = certainly below, Not distinguishable = confidence interval straddles the threshold. Bottom: probability of exceedance (y-axis) for different yield values (x-axis). The vertical black line indicates the threshold value. Left = per cell, center = per province, right = for the whole country.

shows how *Aguila* visualizes these percentiles as exceedance probabilities. The s-shape of the curves indicates that eucalyptus yield is normally distributed at all aggregation levels. The width of the curves, in terms of the range of values they cover, decreases with increasing aggregation level. This is again an indication that uncertainty decreases at higher aggregation levels. By using the option in *Aguila* to show confidence intervals of exceedance probabilities, the percentiles can be used to check, for example, whether a eucalyptus yield of more than 100 kg km⁻² year⁻¹ can be achieved at a certain point in time. The upper part of Fig. 7 is the result of this query. The maps in Fig. 7 show that on a country level (yellow) no definite answer can be given, but at a province level the province Tete (green) can definitely fulfill this condition.

4. Discussion and conclusions

We have shown how a modeler can construct a Spatial Decision Support System (SDSS) that integrates simulation, uncertainty analysis and visualization. This is considered useful in the light that current SDSSs tend to ignore uncertainty (Foody, 2003; Ivanovic & Freer, 2009). The advantages of the constructed PCRaster Land Use Change model (PLUC) are that the uncertainty analysis is coupled to the model, so that the output uncertainty indicators adapt automatically to changes in inputs or parameters, and that the analysis is iterative (Manson, 2007), i.e. evaluated at each time step, which is important in non-linear models. We claim that the output maps and graphs of uncertainty distribution in space and time provide an intuitive way for end users to take uncertainty into account in their decisions. The mode of visualization for uncertain spatio-temporal data was considered suitable by end users without specialist knowledge of statistics in study by Aerts, Clarke, et al. (2003).

Hiemstra and Karssenberg (in press) argue that Monte Carlo results can give a non-expert user the impression that it is hard to make any decision at all, because of the large number of cells that are reported as uncertain. Although it also good to stress the limited extent to which models can answer certain questions about complex systems (Manson, 2007), SDSSs usually do have merits for their end users. We have shown that by providing a means to display for example confidence intervals, with easily understandable qualitative categories *lower*, *higher*, and *not distinguishable* instead of difficult to interpret continuous measures such as full probability distributions, uncertainty information can be used to visualize locations where decisions can be made given a predefined confidence level. A disadvantage is that the error models of the input variables and parameters need to be specified, which is not a straightforward task, especially for users inexperienced in statistics.

Another advantage of the stochastic land use change model is that it provides insights that could have been missed in a deterministic model. For example, where the probability that bioenergy crops could be cultivated is high or low. Running the model at a finer scale on sites with a high potential bioenergy-crop yield, and investigating the effect of actual allocation is the next stage of our research. A different example of the added value of stochastic modeling is the observation of the difference in the shape of concentric rings of land use around cities, which has provided an insight in the combined influence of the range parameters of cities and roads on the resulting land use patterns. This indicates the huge effect that such parameters can have on model outcomes and thereby emphasizes the caution that should be taken in setting such parameters deterministically.

Although the error propagation modeling provides information on the uncertainties in model outputs, it is not a means to evaluate the quality of the internal structure of the model, or to validate the

model by comparing model outputs and independent observational data. Evaluation of the internal model structure would be possible in principle by using detailed observational data on certain sub-processes in the model. However, this data is currently not available for Mozambique. Also, validation is difficult because land use change in Mozambique in the past decades is characterized by civil war, independence and natural disasters, and therefore we do not expect steady continuation of past trends.

A disadvantage for end users of the usage of the Monte Carlo method in an SDSS, also concluded from other studies (Aerts, Goodchild, et al., 2003; Ligmann-Zielinska & Sun, 2010), is computation time. End users do not always have the time to wait at length for their model output. On a desktop pc it took about two days to run the 500 samples. Especially calculation of percentiles is computationally demanding. Another drawback is that PLUC now takes into account data uncertainties and model parameter uncertainties, but not model structure uncertainties, e.g., about the selected error models (Brown & Heuvelink, 2007), model rules, and raster resolution, which are debatable as well. This could be solved by creating a probability distribution over different plausible rules or over the range between coarsest legible resolution and finest legible resolution (Hengl, 2006), so that these can also be sampled in the Monte Carlo simulation. The latter is complex, as spatio-temporal probability distributions alter with changing resolution (Bierkens et al., 2000). Also, having more stochastic parameters complicates model parameterization and raises the number of samples required and thus increases computation time even more. Nevertheless, it is important that uncertainty in simulation models in SDSSs, which grow ever more complex, is somehow evaluated and communicated. This paper shows that this can be accomplished almost without any additional work on the modelers side, which is a major step forward in the exposure of uncertainty in SDSSs.

References

- Aerts, J. C. J. H., Clarke, K. C., & Keuper, A. D. (2003). Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30, 249–261.
- Aerts, J. C. J. H., Goodchild, M. F., & Heuvelink, G. B. M. (2003). Accounting for spatial uncertainty in optimization with spatial decision support systems. *Transactions in GIS*, 7, 211–230.
- Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). A review and assessment of land-use change models: Dynamics of space, time, and human choice, (pp. 1–61). US Department of Agriculture, Forest Service.
- Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Arndt, C., Benfica, R., Tarp, F., Thurlow, J., & Uaiene, R. (2010). Biofuels, poverty, and growth: a computable general equilibrium analysis of Mozambique. *Environment and Development Economics*, 15, 81–105.
- Batidzirai, B., Faaij, A. P. C., & Smeets, E. M. W. (2006). Biomass and bioenergy supply from Mozambique. *Energy for Sustainable Development*, X, 28.
- Batty, M. (2005). Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environment and Planning A*, 37, 1373–1394.
- Bierkens, M. F. P., Finke, P. A., & de Willigen, P. (2000). *Upscaling and Downscaling Methods for Environmental Research*. Dordrecht, The Netherlands: Kluwer Academic Publishers, p. 190.
- Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, 13, 2149–2175.
- Bradshaw, G. A., & Borchers, J. G. (2000). Uncertainty as information: narrowing the science-policy gap. *Conservation Ecology*, 4, <<http://www.consecol.org/vol4/iss1/art7/>>.
- Brown, J. D., & Heuvelink, G. B. M. (2007). The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers and Geosciences*, 33, 172–190.
- Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19, 153–174.
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographical information systems*. Oxford, UK: Oxford University Press.
- Chang, N. B., Parvathinathan, G., & Bredon, J. B. (2008). Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region. *Journal of Environmental Management*, 87, 139–153.

- Chen, H., Wood, M. D., Linstead, C., & Maltby, E. (2011). Uncertainty analysis in a GIS-based multi-criteria analysis tool for river catchment management. *Environmental Modelling and Software*, 26, 395–405.
- Eckhardt, K., Breuer, L., & Frede, H.-G. (2003). Parameter uncertainty and the significance of simulated land use change effects. *Journal of Hydrology*, 273, 164–176.
- ESRI (2010). ArcGIS internet site. <<http://www.esri.com>>.
- Evans, T. P., & Kelley, H. (2004). Multi-scale analysis of a household level agent-based model of landcover change. *Journal of Environmental Management*, 72, 57–72.
- FAO (2003). World agriculture towards 2015/2030 an FAO perspective. In J. Bruinsma (Ed.). Rome: Food and Agriculture Organisation.
- Footy, G. M. (2003). Uncertainty, knowledge discovery and data mining in GIS. *Progress in Physical Geography*, 27, 113–121.
- Geertman, S., & Stillwell, J. (2004). Planning support systems: An inventory of current practice. *Computers, Environment and Urban Systems*, 28, 291–310.
- Goodchild, M. F. (2004). A general framework for error analysis in measurement-based GIS. *Journal of Geographical Systems*, 6, 323–324.
- Goovaerts, P. (2010). Geostatistical software. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 129–138). Heidelberg: Springer.
- Gorsevski, P. V., Gessler, P. E., Boll, J., Elliot, W. J., & Foltz, R. B. (2006). Spatially and temporally distributed modeling of landslide susceptibility. *Geomorphology*, 80, 178–198.
- Haberl, H., Beringer, T., Bhattacharya, S. C., Erb, K. H., & Hoogwijk, M. (2010). The global technical potential of bio-energy in 2050 considering sustainability constraints. *Current Opinion in Environmental Sustainability*, 2, 394–403.
- Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, 32, 1283–1298.
- Heuvelink, G. B. M. (1998). Error propagation in environmental modelling with GIS. Hiemstra, P. H., & Karssenber, D. (in press). Decision support with ensemble forecasting and data assimilation during accidental nuclear releases. Submitted to *Environmental Modelling & Software*.
- Hoogwijk, M., Faaij, A., Eickhout, B., De Vries, B., & Turkenburg, W. (2005). Potential of biomass energy out to 2100, for four IPCC SRES land-use scenarios. *Biomass and Bioenergy*, 29, 225–257.
- INE (2003). Censo Agro-Pecuário 1999–2000, resultados Temáticos. Maputo: Instituto Nacional de Estatística.
- Ivanovic, R. F., & Freer, J. E. (2009). Science versus politics: truth and uncertainty in predictive modelling. *Hydrological Processes*, 23, 2549–2554.
- Jansen, L. J. M., Bagnoli, M., & Focacci, M. (2008). Analysis of land-cover/use change dynamics in Manica Province in Mozambique in a period of transition (1990–2004). *Forest Ecology and Management*, 254, 308–326.
- Karssenber, D., & De Jong, K. (2006). Towards improved solution schemes for Monte Carlo simulation in environmental modeling languages. In P. J. M. Oosterom (Ed.), *Geo-Information and computational geometry*. Delft: NGC (Nederlandse Commissie voor Geodesie, in English: Netherlands Geodetic Commission).
- Karssenber, D., de Jong, K., & van der Kwast, J. (2007). Modelling landscape dynamics with Python. *International Journal of Geographical Information Science*, 21, 483–495.
- Karssenber, D., Schmitz, O., Salamon, P., de Jong, K., & Bierkens, M. F. P. (2010). A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling and Software*, 25, 489–502.
- Kok, K., Farrow, A., Veldkamp, A., & Verburg, P. H. (2001). A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems and Environment*, 85, 223–238.
- Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., et al. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, 11, 261–269.
- Lei, Z., Pijanowski, B. C., Alexandridis, K. T., & Olson, J. (2005). Distributed modeling architecture of a multi-agent-based behavioral economic landscape (MABEL) model. *Simulation*, 81, 503–515.
- Ligmann-Zielinska, A., & Sun, L. (2010). Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change. *International Journal of Geographical Information Science*, 24, 1829–1850.
- Ligtenberg, A., Beulens, A., Kettenis, D., Bregt, A. K., & Wachowicz, M. (2009). Simulating knowledge sharing in spatial planning: an agent-based approach. *Environment and Planning B: Planning and Design*, 36, 644–663.
- Ma, L., Arentze, T., Borgers, A., & Timmermans, H. (2007). Modelling land-use decisions under conditions of uncertainty. *Computers, Environment and Urban Systems*, 31, 461–476.
- Manson, S. M. (2007). Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design*, 34, 245–260.
- NetLogo (2010). NetLogo website. <<http://ccl.northwestern.edu/netlogo/>>.
- Nguyen, T. G., de Kok, J. L., & Titus, M. J. (2007). A new approach to testing an integrated water systems model using qualitative scenarios. *Environmental Modelling and Software*, 22, 1557–1571.
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263, 641–646.
- Pan, Y., Roth, A., Yu, Z., & Doluschitz Reiner, R. (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34, 400–408.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93, 314–337.
- PCRaster (2010). PCRaster internet site. <<http://www.pcraster.geo.uu.nl>>.
- Pebesma, E. J., de Jong, K., & Briggs, D. (2007). Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: An air quality example. *International Journal of Geographical Information Science*, 21, 515–527.
- Pontius, R. G., Jr., & Spencer, J. (2005). Uncertainty in extrapolations of predictive land-change models. *Environment and Planning B: Planning and Design*, 32, 211–230.
- Python (2010). Python programming language. <<http://www.python.org>>.
- Robinson, V. B. (2003). A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS*, 7, 3–30.
- Smeets, E. M. W., Faaij, A. P. C., Lewandowski, I. M., & Turkenburg, W. C. (2007). A bottom-up assessment and review of global bio-energy potentials to 2050. *Progress in Energy and Combustion Science*, 33, 56–106.
- Stella (2010). Stella website. <<http://www.iseesystems.com>>.
- UNDP (2008). World Population Prospects: The 2008 revision. Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.
- van der Hilst, F., Dornburg, V., Sanders, J. P. M., Elbersen, B., Graves, A., Turkenburg, W. C., et al. (2010). Potential, spatial distribution and economic performance of regional biomass chains: The North of the Netherlands as example. *Agricultural Systems*, 103, 403–417.
- van der Hilst, F., Versteegen, J. A., Karssenber, D., & Faaij, A. (submitted for publication). Spatio-temporal land use modelling to assess land availability for energy crops – illustrated for Mozambique. *Global Change Biology Bioenergy*.
- Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: Exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecology*, 24, 1167–1181.
- Verburg, P. H., Schot, P. P., Dijst, M. J., & Veldkamp, A. (2004). Land use change modelling: Current practice and research priorities. *Geojournal*, 61, 309–324.
- von Thünen, J. H. (1966). *Der isolierte Staat in Beziehung auf Landwirtschaft un Nationalökonomie*. Stuttgart: Fischer.
- Watson, H. (2011). Potential to expand sustainable bioenergy from sugarcane in southern Africa. *Energy Policy*, 39, 5746–5750.